دانشگاه صنعتی نوشیروانی بابل

دانشکده مهندسی برق و کامپیوتر

سمینار کارشناسی ارشد

رشته ی مهندسی کامپیوتر گرایش هوش مصنوعی

مروری بر روشهای ارتباط دارو-هدف مبتنی بر شبکه عصبی گراف

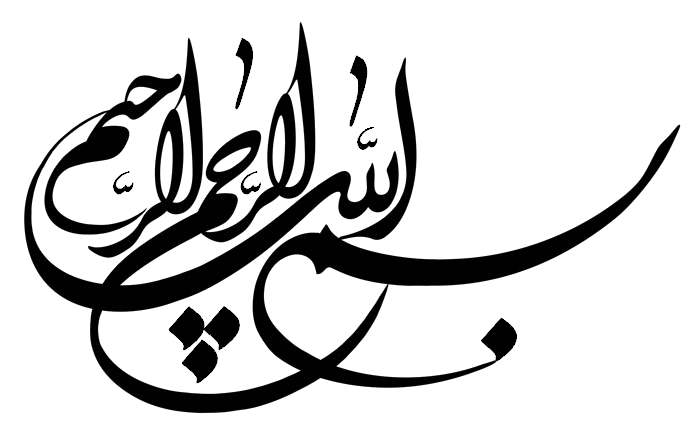
دانشجو:

**سونیا فلاحتی**

استاد راهنما:

**دکتر فاطمه زمانی**

بهار 1402



## چکیده

تولید وتوسعه داروهای جدید پرهزینه و زمان برهستند. چندین رویکرد مهم برای افزایش کارایی در فرآیند کشف و توسعه دارو وجود دارد یکی ازاین رویکردها استفاده مجدد از دارو های موجود است.

استفاده مجدد از داروها به دلیل کاهش هزینه های تحقیق و توسعه ، کاهش زمان تولید دارو، کاهش خطرات جانبی و افزایش اطمینان از اثربخشی دارو ها ، میتواند بسیار موثر و مفید باشد. روشهای یادگیری عمیق، با استفاده از شبکه های عصبی، میتوانند به صورت خودکار و بدون نیاز به دانش خاصی، برای کشف دارو و تحلیل شناسایی ساختار دارو، استفاده شوند. به منظوراستفاده مجدد از داروها به طور موثر، دانستن اینکه کدام پروتئین ها توسط کدام داروها هدف قرار می گیرند مفید است. پیش بینی ارتباط میان دارو-هدف گام مهمی در این راستا می‌باشد. مدل های محاسباتی بسیاری برای پیش بینی ارتباط میان دارو-هدف مطرح شده اند. مدل‌های یادگیری عمیق دارای بهترین عملکردها در پیش‌بینی ارتباط دارو-هدف هستند. تحقیقات انجام شده در یادگیری عمیق معمولا نمایش مولکولهای دارو را به صورت توالی رشته ای در نظر می گیرند که روشی طبیعی برای نشان دادن مولکول ها نیست زیرا هنگام استفاده ازرشته ها، اطلاعات ساختاری مولکول از بین می رود که میتواند قدرت پیش بینی یک مدل و کارایی آن را کاهش دهد.

در سالهای اخیر از شبکه عصبی گراف در پیش بینی ارتباط میان دارو و پروتئین ها استفاده شده است که کارایی بسیار خوبی در مقایسه با سایر روشهای مورد استفاده از خود نشان داده است.

مولکولها را می توان به راحتی با استفاده از یک گراف نشان داد، شبکه عصبی گراف (GNN) برای استخراج ویژگیهای مولکول دارو وپروتئین مناسب است. GNN می تواند اطلاعات ساختاری گراف مولکولی را با استفاده از ویژگی های گره های همسایه به دست آورد. شبکه عصبی گراف می تواند ویژگیهای ساختاری مولکولهارا بهتر به نمایش بگذارد و مستقیماً داده های گراف را پردازش کند.

در این سمینار یک بررسی جامع از کاربردهای اخیر شبکه عصبی گراف در زمینه پیش بینی ارتباط دارو و پروتئین ارائه خواهیم کرد.

**واژه‌های کلیدی:** شبکه عصبی گراف؛ پیش بینی ارتباط دارو-هدف؛یادگیری عمیق؛تولید دارو؛هوش مصنوعی؛استفاده مجدد از دارو.

فهرست مطالب

[چکیده ‌ح](#_Toc123596109)

[فهرست مطالب ‌ط](#_Toc123596110)

[فهرست شکل ‌ها ‌ك](#_Toc123596111)

[فهرست جدول ‌ها ‌ل](#_Toc123596112)

[فصل 1 مقدمه 1](#_Toc123596113)

[فصل 2 مفاهیم اولیه 3](#_Toc123596120)

[فصل 3 مطالعات پیشین 7](#_Toc123596126)

[فصل 4 نتیجه‌گیری 13](#_Toc123596134)

[مراجع 15](#_Toc123596138)

## 

## فهرست شکل‌ها

[شکل ‏3‏‑1: 8](#_Toc123596145)

[شکل پ‏1‏‑‏: 17](#_Toc123596146)

## فهرست جدول‌ها

[جدول ‏3‏‑1: 8](#_Toc123596143)

[جدول پ‏1‏‑: 17](#_Toc123596144)

**فصل 1**

**مقدمه**

هدف اولیه از کشف دارو[[1]](#footnote-1)، تولید داروهای ایمن و موثر برای بیماری های انسان است. فرایند کشف و تولید داروی جدید از مرحله ی شناسایی هدف تا رسیدن به مرحله آزمایشات بالینی بسیار زمان بر و پرهزینه است. به طور معمول حدود 10 تا 15 سال و 0.8 تا 1.5 میلیارد دلارهزینه تولید یک داروی جدید است.

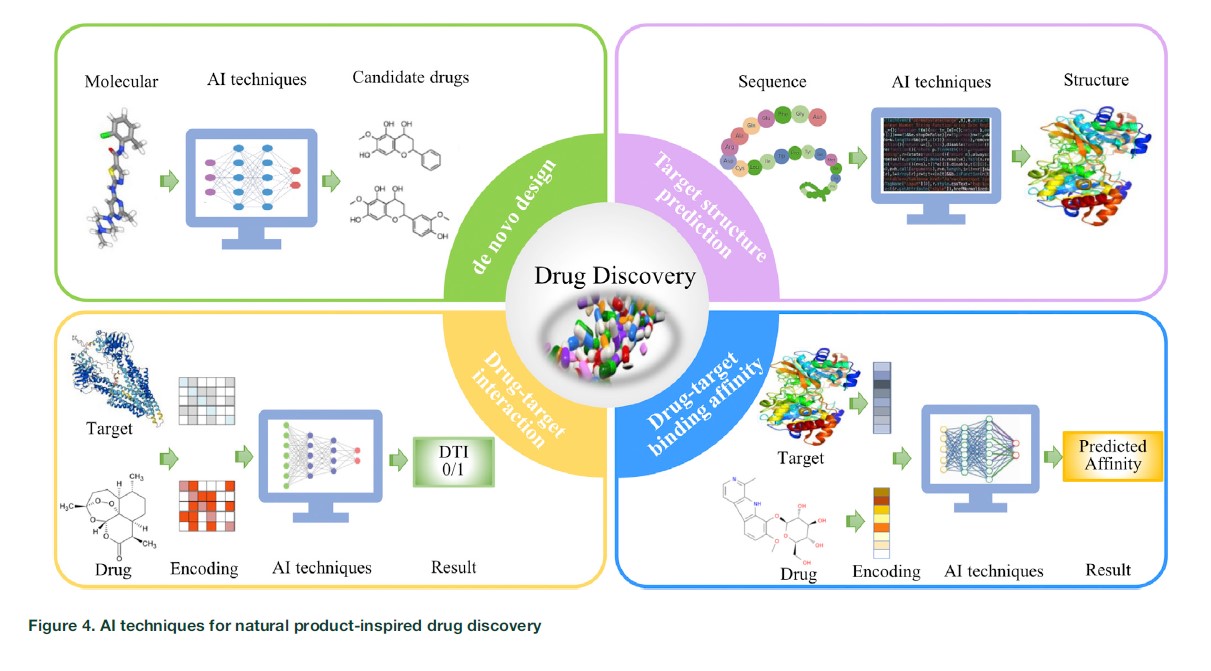
در سالهای اخیر استفاده از هوش مصنوعی[[2]](#footnote-2) (AI) در کشف دارو بسیار مورد توجه قرار گرفته است. هوش مصنوعی در فرایند تولید دارو بهبود قابل توجهی را به همراه دارد. از مزایای استفاده از هوش مصنوعی در فرایند تولید دارو میتوان به کاهش زمان و هزینه های تحقیقات ، کاهش خطا در فرایند تولید، بهبود کیفیت و کارایی دارو، افزایش سرعت و دقت در شناسایی و تولید دارو، پیش بینی خطرات و عوارض جانبی دارو، شناسایی مولکولهای دارویی جدید، شناسایی نقاط ضعف و قدرت مولکولهای دارویی ، بهبود فرآیند طراحی دارو و بهبود فرآیند ارزیابی دارواشاره کرد.

تکنیکهای هوش مصنوعی شامل روشهای مختلفی برای پیش بینی ارتباط میان دارو و پروتئین است. از جمله این روشها میتوان به مدلهای شبکه عصبی، SVM، روشهای رگرسیون خطی و غیرخطی، مدلهای گرافی و غیره اشاره کرد. همچنین از الگوریتمهای ژنتیک، الگوریتمهای تکاملی، الگوریتمهای شبکه عصبی برای پیدا کردن ساختار بهینه دارو و پروتئین استفاده میشود.

استفاده از تکنیک های هوش مصنوعی زمینه ی درک فضاهای بسیار پیچیده بیولوژیکی و استخراج ویژگی ها از داده های ورودیرا فراهم می کند و نتایج چشم گیری را ارائه میدهد. برخی از شرکت های بزرگ داروسازی، مانند Bayer ، Roche و Pfizer نیزفعالیت خود رابرای توسعه روش های مبتنی بر تکنیک های هوش مصنوعی در تولید و کشف دارو آغاز کرده اند. با توجه به نرخ موفقیت پایین، هزینه های بالا و سرعت آهسته کشف و توسعه داروهای جدید، استفاده مجدد از داروهای قدیمی برای درمان بیماری های جدید بسیار مقرون بصرفه است. استفاده مجدد از دارو رویکردی برای تسریع فرآیند کشف دارواست. در واقع به دنبال کشف کاربردهای جدید برای یک داروی موجود هستیم که قبلاً بررسی نشده و در حال حاضر تجویز نشده است.

در فرایند استفاده مجدد از دارو بر اساس داروهای مورد تایید FDA، نیازی به آزمایشات انسانی نیست. همچنین ریسک پایین شکست، چرخه زمانی کوتاهتر، میزان موفقیت بالاتر و سرمایه‌گذاری کمتر از مزایای استفاده مجدد از دارو هستند.

تکنیک های AI تقریبا در تمام جنبه های کشف دارو مورد استفاده قرار گرفته است. در شکل 1 چهار دسته بندی مهم را معرفی و بررسی میکنیم.



شکل 1: تکنیک های هوش مصنوعی در کشف دارو[]

**طراحی و تولید داروهای جدید(De novo drug design)**  : در سالهای اخیر تکنیک های هوش مصنوعی فرصت های جدیدی را برای طراحی و تولید داروهای جدید فراهم آورده اند و روند کشف دارو را تسریع کرده اند. چارچوب اساسی مدل‌های مبتنی بر یادگیری عمیق برای طراحی داروی جدید در سمت چپ بالای شکل 4 نشان داده شده است.

**پیش بینی ساختار هدف (Target structure prediction) :** بیشتر اهداف داروها، پروتئین ها هستند که نقش مهمی در فعالیت آنزیم ها دارند. تکنیک‌های تجربی که برای تعیین ساختار پروتئین ها پیشنهاد شده اند زمان‌بر و پرهزینه هستند.این روشهای قدیمی تنها ساختار 100000 پروتئین منحصر به فرد که تنها بخش کوچکی از پروتئین های شناخته شده را تشکیل می دهند را رمزگشایی کرده‌اند.

با رشد سریع قدرت محاسباتی و پیشرفت های ناشی از تکنیک های هوش مصنوعی، روشهای جدیدی برای پیش بینی ساختار پروتئین ها پیشنهاد شده است که در سمت راست بالای شکل 4 ارائه شده است.

**پیش بینی تعاملات دارو-هدف[[3]](#footnote-3)** (**DTI prediction**) : پیش بینی DTI نقش حیاتی در توسعه دارو ، استفاده مجدد از دارو، و شناسایی عوارض جانبی احتمالی داروها ایفا می کند. به تعاملات میان مولکولهای دارو و پروتئین های هدف اشاره دارد. فرایند پیش بینی DTI در گوشه پایین سمت چپ شکل 4 نشان داده شده است. ابتدا ترکیبات مولکولهای دارو و

پروتئین ها با استفاده از روشهای Embedding کدگذاری می شوند. سپس از آنها به عنوان ورودی در مدلهای مبتنی بر یادگیری عمیق استفاده می شود که به طور قابل توجهی فرایند کشف دارو را تسهیل می کند.

**پیش بینی ارتباط دارو-هدف[[4]](#footnote-4)** (**DTA prediction**) : برای توصیف قدرت اتصال بین یک مولکول دارو و پروتئین استفاده می شود واطلاعاتی در موردارتباط بین یک مولکول دارو و هدف ارائه می دهد. پیش‌بینی دقیق DTA در توسعه اولیه دارو و در طول فرایند استفاده مجدد از دارو بسیار مهم است.

این پیش‌بینی می‌تواند برای تعیین اینکه آیا مولکول دارو می‌تواند به یک هدف متصل شود یا خیر،استفاده شود. روش مبتنی بر توالی رشته ای[[5]](#footnote-5) و روش مبتنی بر ساختار[[6]](#footnote-6) دو روش متداول در پیش‌بینی DTAهستند. در این سمینار به بررسی و مقایسه پژوهش های انجام شده در پیش بینی DTA میپردازیم.

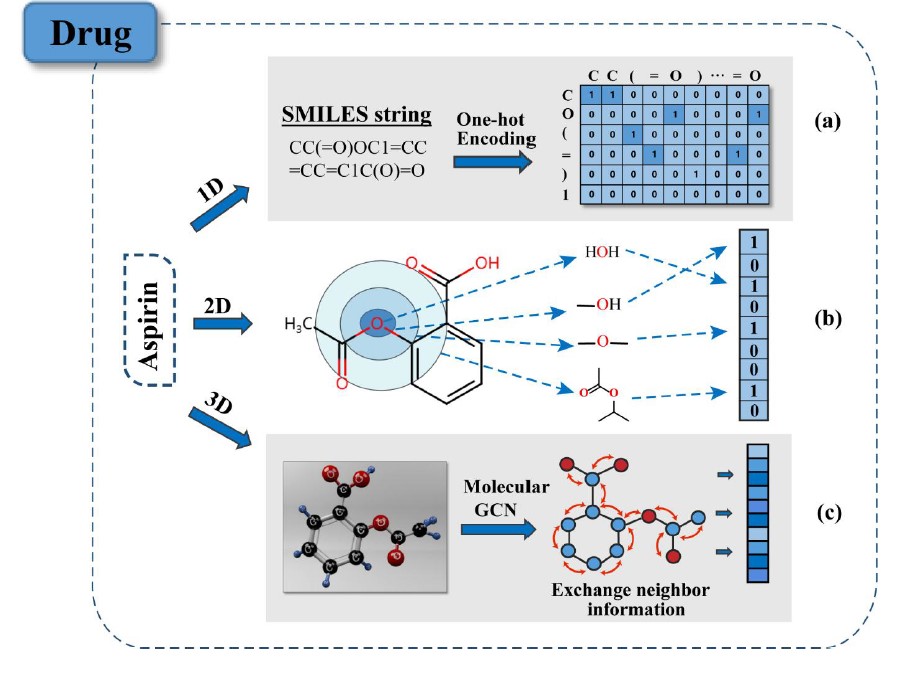
**فصل 2**

**مفاهیم اولیه**

داده های ورودی در مدلهای مبتنی بر DL برای کشف دارو مولکول ها هستند. داروها و اهداف پروتئینی آنها مولکول هایی کوچک یا درشت هستند. برای مشخص کردن این مولکول ها، انواع مختلفی ازنمایش‌های مولکولی استفاده شده است.

بسیاری از روش‌های یادگیری ماشین ML) ) از توالی‌های ساده وویژگی های مولکولی از پیش تعریف شده به صورت دستی استفاده میکنند. از آنجایی که نوع نمایش داده ها بر روی عملکرد مدل در پیش بینی DTA تاثیر بسزایی دارد مطالعات بسیاری در این مورد انجام شده است.

بسیاری از مدل های مبتنی بر DL تمایل دارند از نمایش های سه بعدی استفاده کنند. دارو ها با نمایش های مولکولی مختلفی در شبکه های عصبی عمیق استفاده می شوند که در این بخش معرفی میشوند. انواع مختلف بازنمایی های دارویی بصورت یک بعدی ، دو و سه بعدی که در کشف دارو استفاده می شوند در شکل 2 نشان داده شده است. در این شکل انواع مختلف نمایش داروی آسپرین را مشاهده میکنیم.

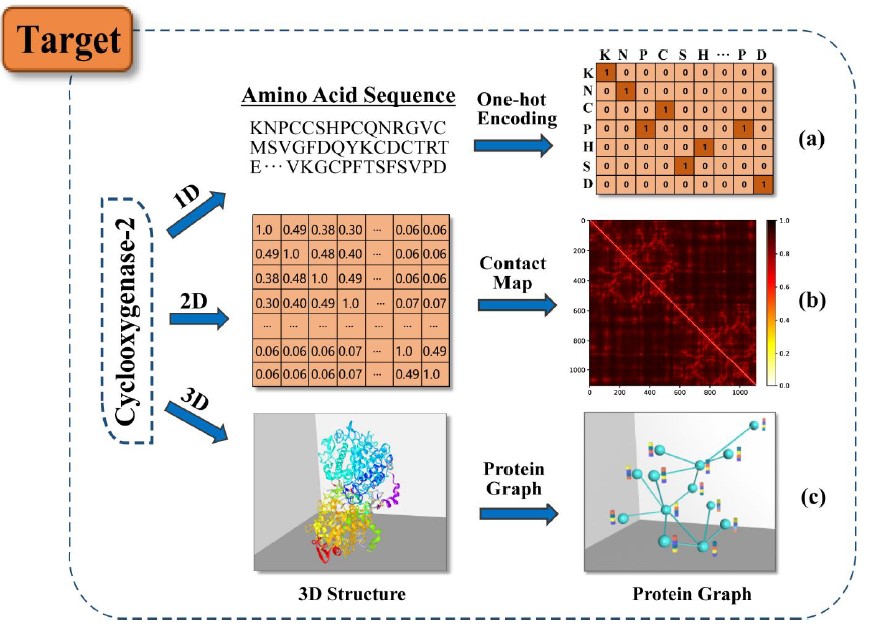


شکل 2 : انواع مختلف نمایش دارو در مدلهای مبتنی بر DL []

SMILES[[7]](#footnote-7) یک نماد متنی برای ترجمه ساختار 3بعدی مولکولهای شیمیایی به صورت رشته است که میتواند برای پردازش در نرم افزار های رایانه ای به کار گرفته شود. رشته های SMILES دارو را می توان به یک گراف مولکولی تبدیل کرد به طوری که اتم‌ها و پیوندها را به‌عنوان گره ها و یال های یک گراف نشان دهیم وبه عنوان ورودی در ساختار GNNها استفاده کنیم.

نمایش پروتئین ها در شکل 3 دیده میشود. رشته های پروتئین از 20 آمینو اسید استاندارد تشکیل شده‌اند که هر کدام از

آنها را می توان به سادگی رمزگذاری کرد.از Map Contact برای نمایش دو بعدی پروتئین استفاده میشود و ساختار 3بعدی آنها بصورت گراف برای پردازش توسط GNN یکار می رود.



شکل 3 : نمایش های مختلف پروتئین در مدلهای مبتنی بر DL []

**معرفی مدل‌های یادگیری عمیق**

مدل های DL را می توان بر اساس هدف، عملکرد، روش یادگیری و ساختار آنها طبقه بندی کرد.معمولا در توسعه دارو از ترکیب دو یا چند مدل برای دستیابی به نتایج بهتر استفاده میشود. انواع مختلفی از مدل های DL وجود دارد، اما فقط تعدادی از آنها در این بخش توضیح داده میشود. ما نقاط قوت و ضعف هر مدل را شرح می دهیم و مشخصات این مدل ها را در فرایند کشف دارو معرفی میکنیم.

1**- پرسپترون‌های چندلایه[[8]](#footnote-8) (MLP)** :

پرسپترون چند لایه رایج ترین ساختار شبکه عصبی است MLPها از یک لایه ورودی و یک لایه خروجی تشکیل شده‌اند که تمامی آن‌ها به هم متصل شده‌اند. این شبکه‌ها دارای تعداد لایه ورودی و خروجی یکسانی هستند اما ممکن است چندین لایه مخفی داشته باشند.

2**-شبکه‌های عصبی کانولوشنی : (CNN)[[9]](#footnote-9)**

CNNها چندین لایه دارند که ویژگی‌ها را پردازش و از داده‌ها استخراج می‌کنند و به طور گسترده‌ برای شناسایی تصاویر ماهواره‌ای، پردازش تصاویر پزشکی، پیش بینی سری‌های زمانی و تشخیص ناهنجاری‌ها مورد استفاده قرار می‌گیرند.

**3-شبکه‌های حافظه طولانی کوتاه مدت: (LSTM) [[10]](#footnote-10)**

نوعی شبکه عصبی بازگشتی هستند که می‌توانند وابستگی‌های طولانی مدت را یاد بگیرند و به خاطر بسپارند.این شبکه‌ها اطلاعات را در طول زمان حفظ می‌کنند و در پیش بینی سری‌های زمانی مفید هستند زیرا ورودی‌های قبلی را به خاطر می‌آورند. LSTM ‌ها دارای یک ساختار زنجیره‌ای هستند که در آن، چهار لایه در تعامل با هم، به روشی منحصر به فرد ارتباط برقرار می‌کنند. LSTM‌ها معمولاً برای تشخیص گفتار و توسعه دارو نیز استفاده می‌شوند.

**4-شبکه‌های عصبی بازگشتی[[11]](#footnote-11) : (RNN)**

 RNNها، دارای اتصالاتی هستند که حلقه‌هایی جهت­‌دار را تشکیل می‌دهند، که به خروجی LSTM اجازه می‌دهد تا به عنوان ورودی به فاز فعلی تغذیه شود. خروجی LSTM، ورودی مرحله فعلی می‌شود و می‌تواند ورودی‌های قبلی را با توجه به حافظه داخلی خود حفظ کند . RNN ‌ها معمولاً برای زیرنویس تصاویر، تجزیه و تحلیل سری‌های زمانی، پردازش زبان طبیعی، تشخیص دست خط و ترجمه ماشین و و توسعه دارو استفاده می‌شوند.

**5- شبکه‌های مولد متخاصم[[12]](#footnote-12) : (GAN)**

GAN‌ها، از مدلهای  یادگیری عمیق  مولد هستند که نمونه‌های جدیدی از داده‌ها را شبیه به داده‌های آموزشی ایجاد می‌کنند. GAN دو بخش اصلی دارد:

یک مولد (generator) که یاد می‌گیرد داده‌های جعلی تولید کند و یک متمایزگر یا تشخیص دهنده (discriminator)، که از این اطلاعات نادرست درس می‌گیرد.

استفاده از GAN‌ها طی یک دوره زمانی بسیار افزایش یافته است. توسعه دهندگان بازی‌های ویدئویی، از GAN برای ارتقای بافت‌های با وضوح پایین و دو بعدی در بازی‌های ویدئویی قدیمی با بازآفرینی آن‌ها در وضوح 4K یا بالاتر، از طریق آموزش تصاویر استفاده می‌کنند. GAN‌ها به تولید تصاویر واقعی و شخصیت‌های کارتونی، ایجاد عکس از چهره انسان و ارائه اشیاء سه بعدی کمک می‌کنند.

**6- اتوانکدرها[[13]](#footnote-13) یا خودرمزگذارها :**

خودرمزگذارها ([Autoencoders](https://www.simplilearn.com/other-forms-of-deep-learning-tutorial" \t "_blank))، نوع خاصی از شبکه عصبی پیشخور هستند که در آن‌ها، ورودی و خروجی یکسان است. خودرمزگذارها شبکه‌های عصبی آموزش دیده‌ای هستند که داده‌ها را از لایه ورودی به خروجی همانندسازی (replicate) می‌کنند. از خودرمزگذارها برای اهدافی مانند اکتشافات دارویی، پیش بینی محبوبیت و پردازش تصویر استفاده می‌شود. یک خودرمزگذار از سه قسمت اصلی تشکیل شده است: رمزگذار، رمز و رمزگشا.

خودرمزگذارها به گونه‌ای ساختار یافته‌اند تا ورودی را دریافت کرده و آن را به یک بازنمایی متفاوت تبدیل کنند. سپس سعی می‌کنند ورودی اصلی را تا آنجا که ممکن است، به درستی بازسازی کنند.

**7- شبکه عصبی گراف (GNN) [[14]](#footnote-14):**

گراف‌ها ساختارهای داده‌ای قوی هستند که شامل روابط بین اشیا هستند.کاربرد های زیادی در بینایی ماشین وپردازش زبان طبیعی و پزشکی و نوسعه دارو دارند. GNN‌هابرای انجام وظایف پیش‌بینی در سطح گره، سطح لبه و سطح گراف مفید هستند.گره‌ها و یال ‌های گراف ساختار یک مولکول را نشان می‌دهند. گره ها اتم هستند و یال ها پیوندهای شیمیایی هستند. از گراف برای تجزیه و تحلیل ساختار مولکول‌ها استفاده می شود.

**8-مکانیسم توجه[[15]](#footnote-15) در یادگیری عمیق :**

مکانیسم توجه در یادگیری عمیق تکنیکی است که برای بهبود عملکرد یک شبکه عصبی با اجازه دادن به مدل برای تمرکز روی مهم‌ترین داده‌های ورودی در حین تولید پیش‌بینی استفاده می‌شود.این کار با وزن دادن به داده های ورودی انجام می شود تا مدل برخی از ویژگی های ورودی را بر سایرین اولویت دهد. در نتیجه، مدل می‌تواند پیش‌بینی‌های دقیق‌تری را تنها با در نظر گرفتن مهم‌ترین متغیرهای ورودی تولید کند. مکانیسم توجه اغلب در کارهای پردازش زبان طبیعی مانند ترجمه ماشینی استفاده می شود، جایی که مدل باید به بخش های مختلف عبارت ورودی توجه کند تا معنای آن را به طور کامل درک کند و ترجمه مناسب ارائه دهد.

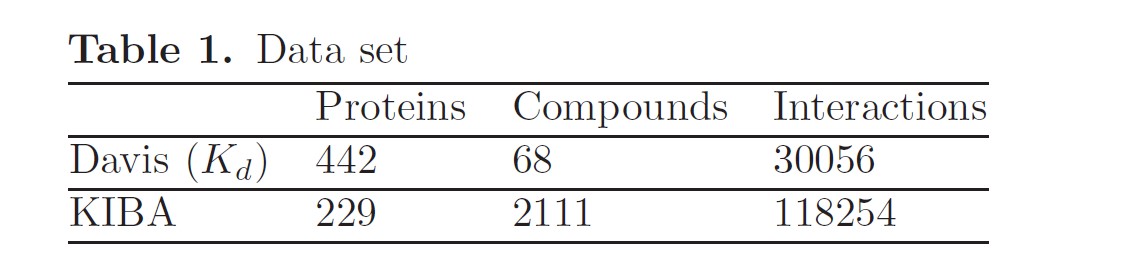
**مجموعه داده** : مجموعه داده های معیاری که برای ارزیابی عملکرد مدل پیش بینی DTA استفاده می شود شامل دو مجموعه داده Davis و KIBAاست که در اکثر مقالات مورد استفاده قرار گرفته است.در ادامه بطور مختصر هر دو مجموعه داده معرفی میشوند.

**معرفی مجموعه داده ها :**

Davis یک مجموعه داده از Kinase است که توسط (دیویس و همکاران سال2011) معرفی شد شامل 25046 جفت دارو- هدف است.

مجموعه داده Kiba توسط (تانگ و همکاران سال2014) ارائه شده است. رشته هایSMILES ازهر دو مجموعه داده از پایگاه داده PubChem و رشته های پروتئین از پایگاه داده پروتئین UniProt استخراج شده است.

مشخصات هردو مجموعه داده در جدول زیر آورده شده است:



جدول 1 : معرفی مجموعه داده

**معیار های ارزیابی مدل پیشنهادی**

برای ارزیابی مدل از concordance index(CI) و mean square error (MSE) استفاده میشود.

CI میزان احتمال شباهت بین مقادیر واقعی و پیش بینی شده را اندازه گیری می کند، فرض کنید δi و bi به ترتیب بیانگر مقدار واقعی و مقدار پیش‌بینی‌شده نمونه i باشند. بنابراین CI به صورت زیر تعریف می شود:

(1)

MSE نشان دهنده تفاوت بین مقادیر پیش بینی شده و مقادیر واقعی است در طول آموزش، یک مدل تلاش می کند تا اختلاف بین مقدار واقعی و پیش بینی شده را کاهش دهد.

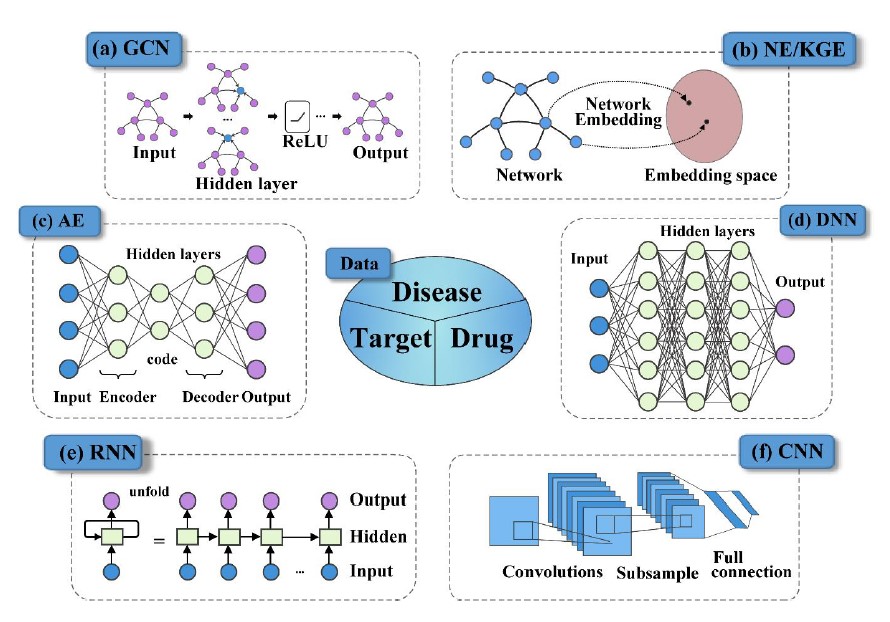
(MSE) به عنوان loss function استفاده میشود، در MSE ، P بردار پیش بینی است و Y مربوط به بردار خروجی های واقعی است و n تعداد نمونه ها را نشان می دهد:

(2)

**فصل3**

**مرورمطالعات پیشین**

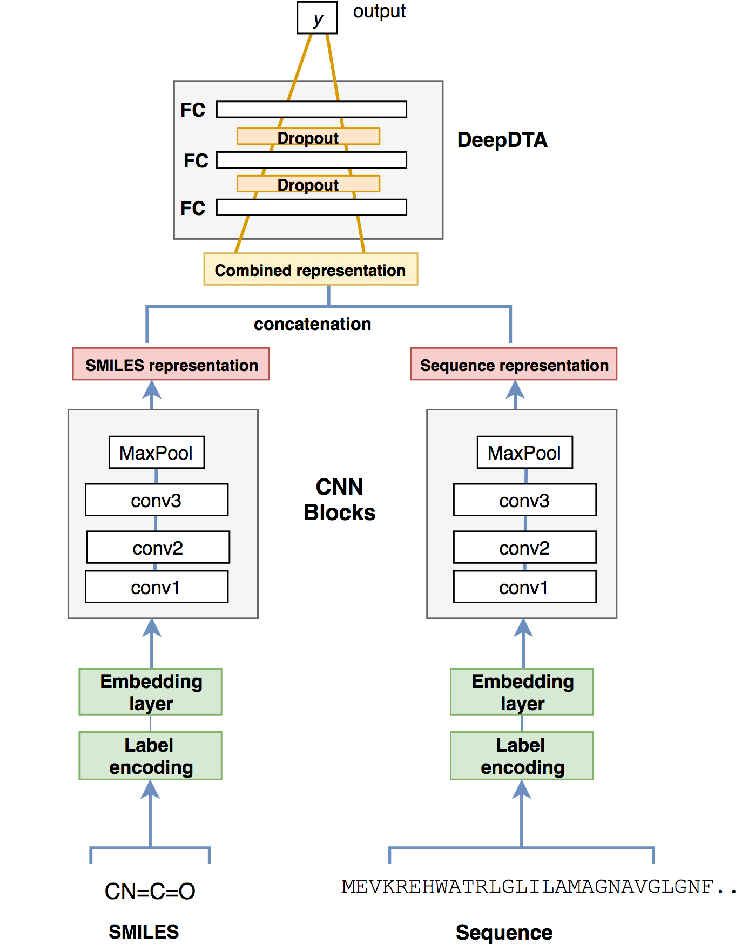
در روشهای یادگیری عمیق پیش بینی ارتباط میان دارو-هدف بر اساس خواص مولکولی دارو و رشته های پروتئین انجام میشود .معماری های مختلفی از این تکنیک ها در شکل3 مشاهده میشود.

****

شکل 3: معماری های مختلف یادگیری عمیق در استفاده مجدد دارو[]

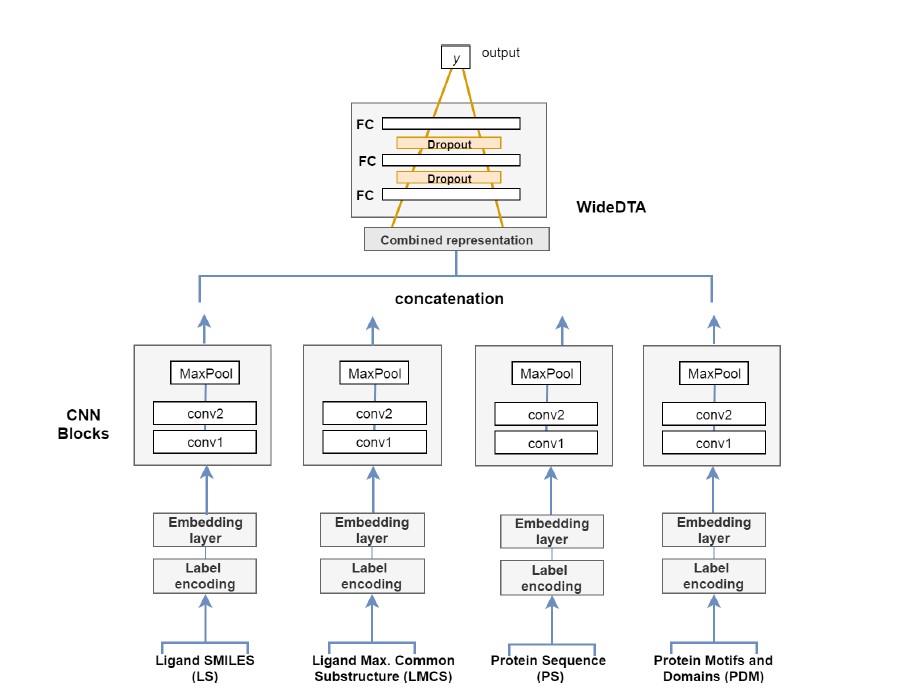
لی و همکاران در سال2015 یک روش مبتنی بر جنگل تصادفی[[16]](#footnote-16) را پیشنهاد کردند[4]. این مدل RF همان سال در KronRLS ، نیز مورد استفاده قرار گرفت که از similarity matrixبرای بهبود عملکرد پیش‌بینی DTA استفاده کرد[5].در سال 2017مدل SimBoost مطرح شد که یک بردار ویژگی را با هر جفت دارو-هدف مرتبط می کند از جفت‌هایی با ویژگی‌های اتصال مشاهده‌شده، یک مدل رابرای یادگیری آموزش می‌دهد[6].

در سال 2018 مدل DeepDTA [7] معرفی شد ، یک مدل مبتنی بر یادگیری عمیق است که از شبکه های عصبی کانولوشنCNN) [[17]](#footnote-17)) برای یادگیری و نمایش رشته های پروتئین و رشته های SMILES استفاده کرد. DeepDTA بهتر از KronRLS و الگوریتم های SimBoost عمل کرده است و نمایش‌های حاصل از بلوک های CNN راباهم ترکیب کرده تا به یک لایه کاملاً متصل وارد شوند و میل ترکیبی دارو-هدف را پیش‌بینی کند.

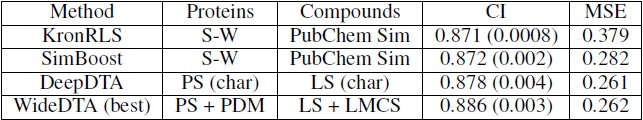


شکل 4 : مدل DeepDTA پیشنهاد شده در سال 2018[7]

WideDTA [8]در سال 2019 معرفی شده است توسعه ای از DeepDTA است. از CNN برای یادگیری بازنمایی داروها و پروتئین ها استفاده می کند. در WideDTAتوالی SMILES و پروتئین به‌عنوان کلمات نظرگرفته شده اند و به ترتیب با یک دنباله هشت کاراکتری و یک دنباله سه کاراکتری ، نشان داده شدند. این مدل چهار قطعه متنی مختلف از اطلاعات مربوط به پروتئین ها و لیگاندها را ترکیب می کند.



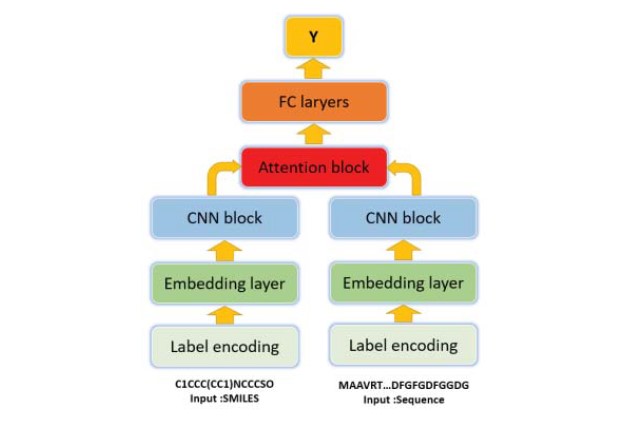
شکل 5 : مدل پیشنهادی WideDTAدر سال 2019[8]



جدول 1: مقایسه نتایج حاصل ازWideDTAو سایر مدلهای مطرح شده

DeepAffinity []یک مدل یادگیری عمیق نیمه نظارت‌شده که ازترکیب RNN وCNN برای پیش‌بینی DTA استفاده می‌کند. از داده‌های بدون برچسب و برچسب‌گذاری‌شده، برای رمزگذاری مشترک مولکولهای دارو و پروتئین وهمینطور از مکانیسم‌ توجه برای تمرکز برروی بخشهای مهم پروتئین و دارو استفاده میکند تا پیش بینی دقیقتری را ارائه دهد.

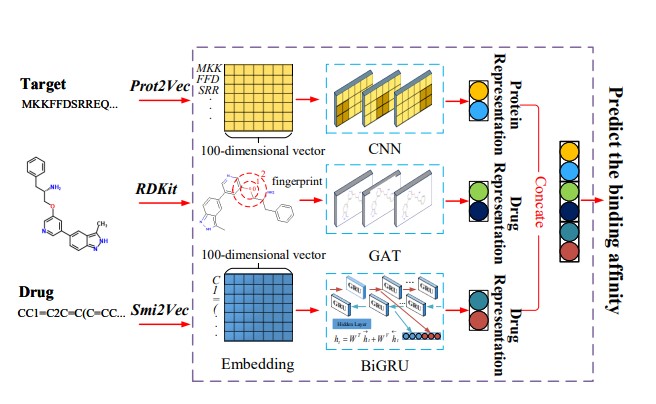
مدل مبتنی بر یادگیری عمیق دیگری ، به نام AttentionDTA [] ، که از مکانیسم توجه برای پیش‌بینی DTA استفاده می‌کند. متفاوت از مدل‌هایی است که از ساختارهای سه‌بعدی برای دارو-هدف یا نمایش گرافی داروها و پروتئین‌ها استفاده می‌کنند، با استفاده از مکانیسم توجه روی بخشهایی از رشته های دارو و پروتئین که هنگام پیش‌بینی میل ترکیبی آنها مهم هستند،تمرکز میکند. دو بلوک کانولوشن 3 لایه جداگانه برای بازنمایی دارو و پروتئین بکار گرفته و سپس یک بلوک توجه در ادامه قرار گرفته است. در مقایسه با DeepDTA نتایج نشان می دهد که مدل مبتنی بر توجه می تواند به طور موثرتری اطلاعات و ویژگیهای دارو و پروتئین را استخراج کند.



شکل7: معماری مدل AttentionDTA []

GANsDTA [9]یک مدل نیمه نظارت شده[[18]](#footnote-18)جدید مبتنی بر شبکه‌های متخاصم مولد (GANs)[[19]](#footnote-19) برای پیش‌بینی DTA بین رشته های SMILES و پروتئین پیشنهاد کرد. مدل GANsDTA می‌تواند با استفاده از دو GAN بر روی داده‌های برچسب‌دار و بدون برچسب، ویژگی‌های پروتئین‌ها و داروها را استخراج کند.

. [10]DeepGS روش دیگری است که از SMILES و رشته های پروتئین برای پیش‌بینی DTA استفاده می‌کند. از GAT برای استخراج ویژگیهای گراف دارو و ازBiGRUبرای دستیابی به اطلاعات شیمیایی آنها استفاده میکند.با تکنیکهای Embedding برای رمزگداری رشته های SMILES و آمینو اسید های پروتئین استفاده میکند.



شکل7: معماری مدل DeepGS []

این مدلهای یادگیری عمیق داروها را به شکل string نمایش میدادند که با ساختار طبیعی مولکولها متفاوت است. این روش‌ها نشان داده‌اند که شبکه‌های عمیق بهتر از الگوریتم‌های یادگیری ماشین سنتی می‌توانند عمل کنند.

هنگام استفاده ازرشته ها، اطلاعات ساختاری مولکولها از بین می رود ، که می تواند قدرت پیش بینی یک مدل و همچنین عملکرد آنرا مختل کند.

در سالهای اخیر شبکه عصبی گراف (GNN) به طور گسترده در زمینه های مختلف استفاده شده است. ازیک گراف متشکل از گره ها و یال ها به عنوان ورودی GNN استفاده می شود و هیچ محدودیتی برای اندازه گراف ورودی وجود ندارد ، که قابلیتی انعطاف پذیر برای استخراج اطلاعات مولکول ها را ارائه می دهد.

شبکه کانولوشن گراف[[20]](#footnote-20) (GCN) و شبکه توجه گراف[[21]](#footnote-21) (GAT) مدل‌های پرکاربرد GNN هستند و به تدریج در طراحی دارو ، مانند پیش‌بینی خواص دارو و همینطور پیش بینی DTA استفاده شده‌اند.

DTA [12]Graph- در سال 2020 نمایش گراف را برای استفاده از اطلاعات ساختار مولکولها ی دارو معرفی کرد و از GNN برای پیش بینی DTA استفاده کرده است. نتایج تأیید می‌کند که در مدل‌های یادگیری عمیق برای پیش‌بینی DTA ، نمایش داروها به صورت گراف می تواند منجر به بهبود عملکرد مدل شود. این مدل ترکیبات دارویی را به عنوان گرافی از فعل و انفعالات بین اتم در نظر گرفته و مدل را حول این مفهوم پیاده سازی میکند. برای توصیف یک گره در گراف، از مجموعه ای از ویژگی های اتمی استخراج شده از DeepChem (رامسوندار و همکاران، 2019) استفاده می کند. هر گره ، یک بردار ویژگی باینری چند بعدی است که شامل پنج مورد اطلاعات زیر است:

نماد اتم،

تعداد اتم های مجاور،

تعداد هیدروژن های مجاور،

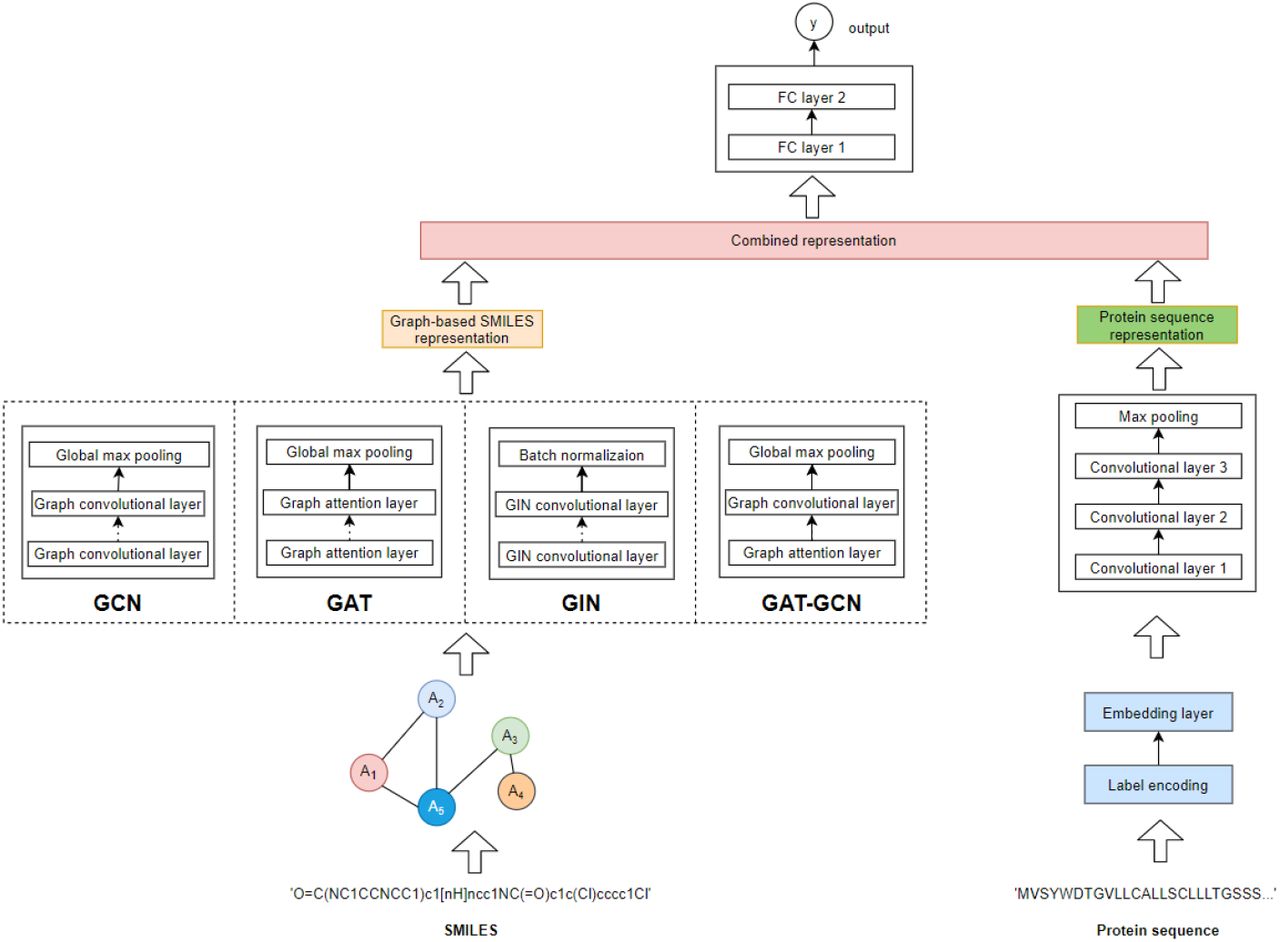
implicit value اتم

و aromatic structure**.**

در واقع کد SMILES را بااستفاده از RDKit و استخراج ویژگی های اتمی به گراف مولکولی متناظرش تبدیل میکند.یک مدل پیش‌بینی DTA را بر اساس ترکیبی از شبکه های عصبی گراف و CNN پیشنهاد می‌کند. برای پروتئین ها از رشته ای ازکاراکترهای ASCII استفاده می کند و چندین لایه CNN 1 بعدی را روی آنها اعمال میکند.

برای دارو از 4مدل مختلف GNNشاملGCN(Graph convolutional network)،

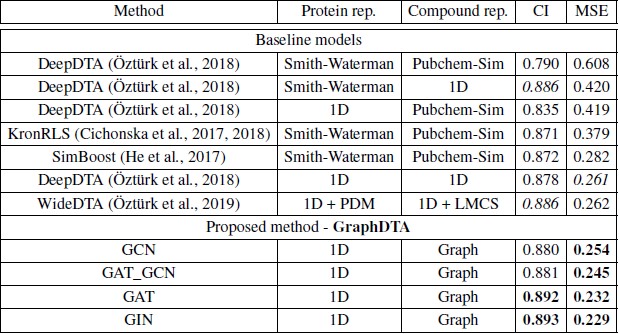
GAT ، ( (graph attention network،) graph isomorphic network)GIN و ترکیبی از GAT-GCN استفاده میکند.

****

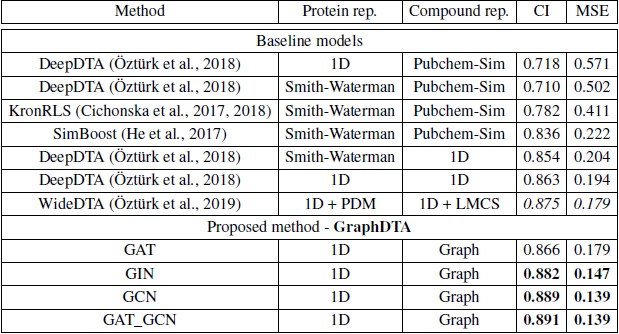
شکل7: معماری مدل پیشنهاد شده GraphDTA [12]

GraphDTA در 3 مرحله کار می کند : ابتدا کد SMILESیک دارو به یک گراف مولکولی تبدیل می‌شود و یک الگوریتم یادگیری عمیق، نمایش گراف را می‌آموزد. در همین حال، رشته پروتئین کدگذاری و تعبیه شده است و چندین لایه کانولوشن 1 بعدی نمایش آن را یاد می گیرند. در نهایت، دو بردار نمایش با هم ادغام میشوند و از چندین لایه کاملاً متصل عبور می کنند تا خروجی را تخمین بزنند.

نتایج حاصل از این مدل روی مجموعه داده های Kiba وDavis در جدول 2 و3 مشاهده می شود.



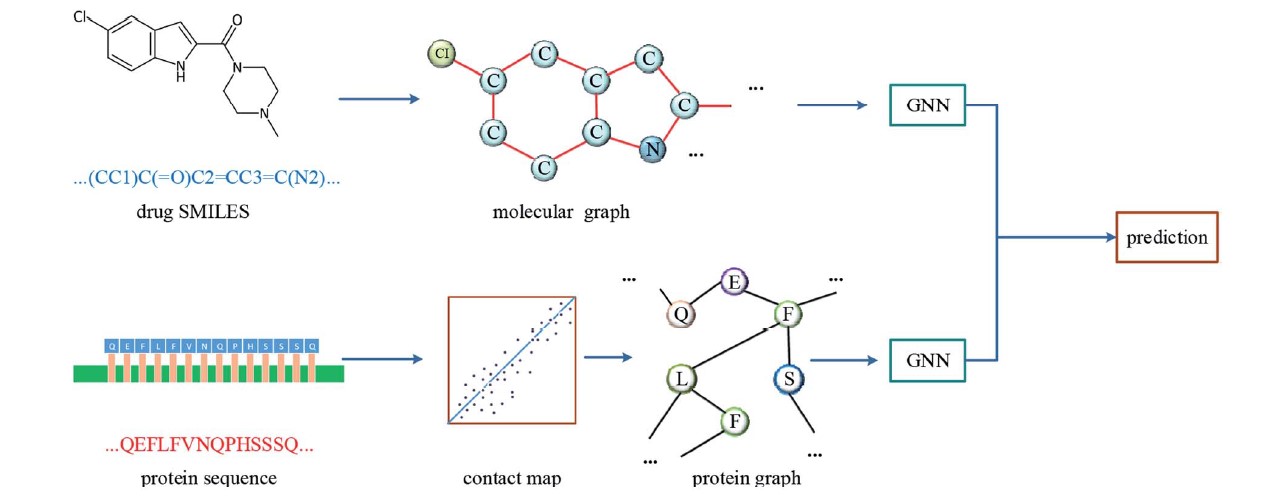
جدول2:نتایج مدل Graph-DTA روی دیتاست Davis در مقایسه با سایر مدل ها[12]



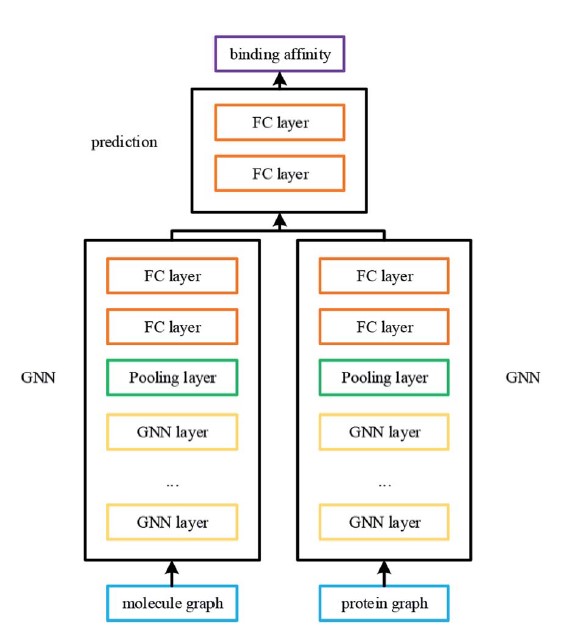
جدول 3 : نتایج مدلGraph-DTA روی دیتاست Kiba در مقایسه با سایر مدل ها[12]

نتایج نشان می دهد که GraphDTA نسبت به سایر مدلهای مطرح شده عملکرد بهتری در پیش بینی DTAداشته است. اما چالشی که مطرح میشود این است که در GraphDTA، CNN برای به دست آوردن ویژگی‌های پروتئینی از طریق رشته ها مورد استفاده قرار گرفت و گرافی برای هر پروتئین ایجاد نشد.

DGraphDTA [13] مدلی است که علاوه بر دارو،پروتئین ها را هم به صورت گراف در نظر میگیرد و برای پیش بینی DTA پیشنهاد شده است. با استفاده از GNN، DGraphDTA می تواند رشته های پروتئین رابر اساس contact map به گراف تبدیل کند و می‌تواند ویژگی‌های ساختاری پروتئین را با توجه به رشته های آنها پیش‌بینی کند.

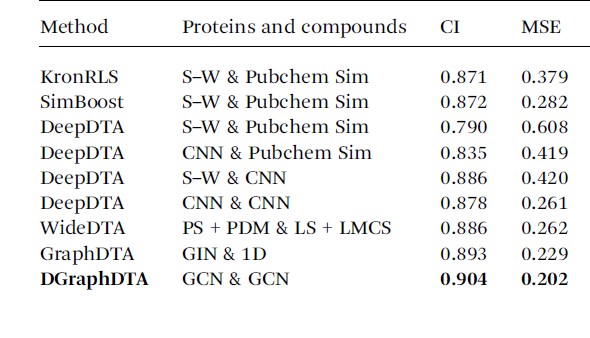


شکل8 : نمایش دارو و پروتئین در مدل پیشنهادی [13]DGraphDTA



شکل9 : معماری مدل پیشنهادی [13] DGraphDTA

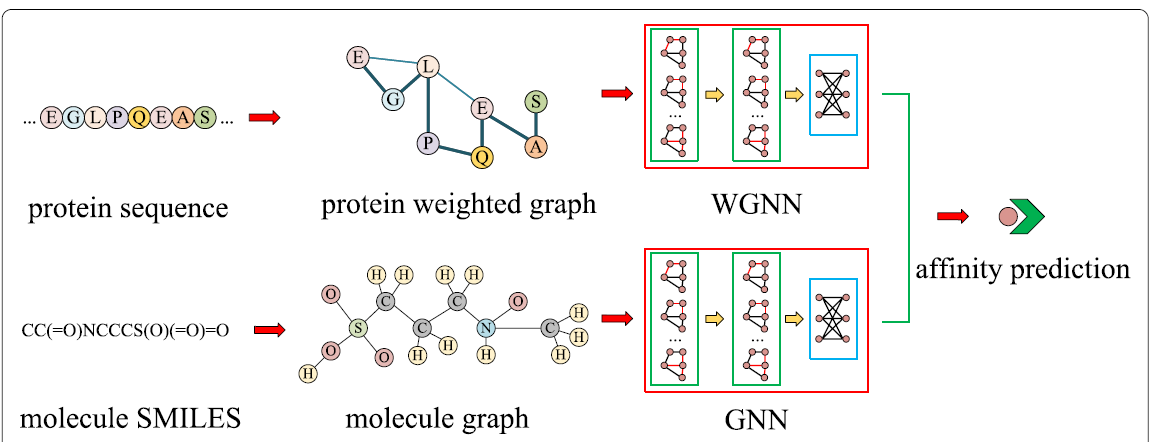
همانطور که درجدول 4 مشاهده می شود در مقایسه با DeepDTA، WideDTA و GraphDTA این مدل پیشنهادی با GCN های سه لایه بهبود عملکرد قابل توجهی دارد.



جدول 4 : مقایسه نتایج مدل [13] DGraphDTA

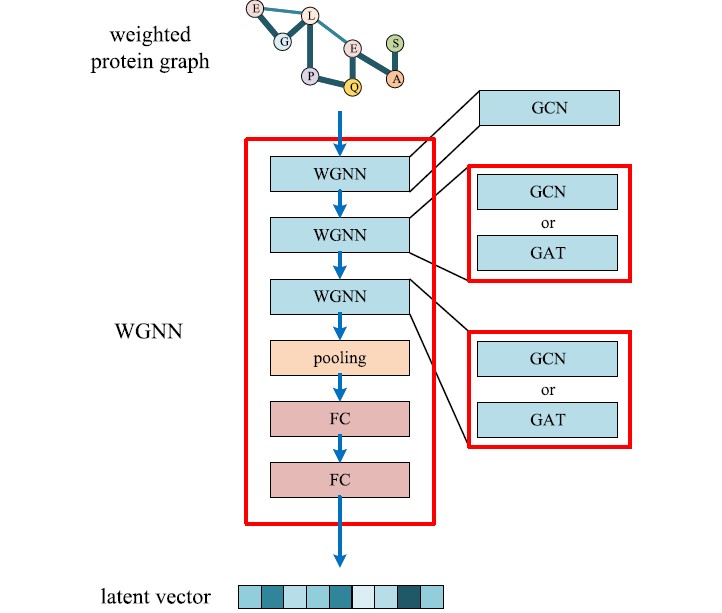
WGNN‑DTA [14] یک مدل گراف وزندار است که می تواند برای پروتئین ها گراف وزندار مناسبی ایجاد کند و آزمایش های مختلف نشان می دهد این مدل برای پیش بینی DTA سریعتر از مدل DGraphDTA است و دقت بالایی دارد.

ورودی ها ی مدل رشته پروتئین و رشتهSMILES است. یک گراف وزنی برای پروتئین ساخته شده است که می تواند به طور جامع اطلاعات ساختار پروتئین را شرح دهد و به جای مقدار باینری که در DGraphDTA استفاده می شود، از احتمال تعامل به عنوان وزن یال های گراف استفاده می شود.



شکل10 : مدل WGNN-DTA ] 14 [

از دو نوع GNN، (GCN) و (GAT) برای استخراج بردارهای نهفته گراف های پروتئین و دارو ، و پیش‌بینی DTAبر اساس بردارهای نهفته استفاده میکند.



شکل10 : معماری مدل WGNN-DTA برای گراف وزندار پروتئین ] 14 [

در جدول زیر خلاصه ای از پژوهش های انجام شده در پیش بینی DTA ذکر شده است.

|  |  |  |
| --- | --- | --- |
| مقاله | روش | توضیحات |
| KronRLS  [2015] | از جنگل تصادفی و  Similarity matrix برای پیش بینی DTA استفاده کرد. | اولین مدل برای پیش‌بینی DTA بدون استفاده از داده‌های ساختاری سه‌بعدی است. یک روش مبتنی بر ML است که از ماتریس‌های شباهت هدف-هدف و شباهت دارو-دارو به عنوان ویژگی برای پیش‌بینی مقادیرDTA استفاده می‌کند. |
| SimBoost  [2017] | یک روش ماشین لرنینگ است  از gradient boosting  برای پیش بینی DTAاستفاده میکند. | این مدل از ماتریس های شباهت و ویژگی های ساخته شده به دست آمده از شباهت هدف-هدف، شباهت دارو-دارو، و شبکه های دارو-هدف برای پیش بینی DTA استفاده میکند.  SimBoost عملکرد پیش بینی را در مقایسه با روش KronRLS بهبود بخشید. |
| DeepDTA  [2018] | یک مدل مبتنی بر یادگیری عمیق است که از شبکه های عصبی کانولوشنCNN)) استفاده کرد. | یکی از اولین روش‌های پیش‌بینی DTA است که از دو CNN اعمال شده بر روی رشته های SMILES دارو و توالی‌های آمینو اسیدی پروتئین ها برای پیش بینی DTAاستفاده میکند. DeepDTA بهتر از KronRLS و الگوریتم های SimBoost عمل کرده است. |
| WideDTA  [2019] | یک مدل مبتنی بر یادگیری عمیق است از اطلاعات متنی و بیولوژیکی دارو و هدف برای پیش‌بینی DTA استفاده می‌کند. | مدل WideDTA توسعه ای از DeepDTA است. از CNN برای بازنمایی داروها و پروتئین ها استفاده می کند. در WideDTAرشته های SMILES و پروتئین به‌عنوان کلمات نظرگرفته شده اند و به ترتیب با یک دنباله هشت کاراکتری و یک دنباله سه کاراکتری ، نشان داده شدند.  WideDTA چهار قطعه متنی مختلف از اطلاعات مربوط به پروتئین ها و لیگاندها را ترکیب می کند. |
| DeepAffinity  [2019] | یک مدل یادگیری عمیق نیمه نظارت‌شده که ازترکیب RNN و CNN برای پیش‌بینی DTA استفاده می‌کند. | از داده‌های بدون برچسب و برچسب‌گذاری‌شده، برای رمزگذاری مشترک مولکولهای دارو و پروتئین وهمینطور از مکانیسم‌ توجه برای تمرکز برروی بخشهای مهم پروتئین و دارو استفاده میکند. |
| AttentionDTA  [2019] | مدل مبتنی بر یادگیری عمیق، که از مکانیسم توجه برای پیش‌بینی DTA استفاده می‌کند. | نتایج نشان می‌دهد که مدل مبتنی بر توجه می‌تواند به طور موثر ویژگی‌های پروتئین مربوط به اطلاعات دارویی و ویژگی‌های دارویی مرتبط با اطلاعات پروتئین را برای پیش‌بینی DTA استخراج کند. |
| GANsDTA  [2020] | مدل نیمه نظارت شده جدید مبتنی بر شبکه‌های متخاصم مولد (GANs) برای پیش‌بینی DTA است. | مدل GANsDTA می‌تواند با استفاده از دو GAN بر روی داده‌های برچسب‌دار و بدون برچسب، ویژگی‌های پروتئین‌ها و داروها را استخراج کند. |
| DeepGS  [2020] | مدل مبتنی بر یادگیری عمیق از ترکیب CNN و GATو بلوک های BiGRU استفاده میکند. | از GAT برای استخراج ویژگیهای گراف دارو و ازBiGRUبرای دستیابی به اطلاعات شیمیایی آنها استفاده میکند.با تکنیکهای Embedding برای رمزگداری رشته های SMILES و آمینو اسید های پروتئین استفاده میکند. |
| GraphDTA  [2020] | نمایش گراف را برای مولکولها دارو معرفی کرد و از ترکیب انواع GNN و CNN برای پیش بینی DTA استفاده کرده است. | تبدیل رشته های SMILES دارو به گراف می تواند منجر به بهبود عملکرد مدل شود. GraphDTA از چهار نوع شبکه عصبی گراف از جمله GCN، GAT، GIN، و یک معماری ترکیبی GAT-GCN، برای پیش بینی DTA استفاده کرده است.  نتایج نشان می دهد که GraphDTA نسبت به سایر مدلهای مطرح شده عملکرد بهتری در پیش بینی DTAداشته است. |
| DGraphDTA  [2020] | با استفاده از GNNهای مختلف روی گراف دارو و گراف پروتئین کار میکند. | مدلی است که علاوه بر دارو،پروتئین ها را هم به صورت گراف در نظر میگیرد.با استفاده از GNN، DGraphDTA می تواند رشته های پروتئین رابر اساس contact map به گراف تبدیل کند و می‌تواند ویژگی‌های ساختاری پروتئین را با توجه به رشته های آنها پیش‌بینی کند. |
| WGNN-DTA  [2022] | با استفاده از GNNهای مختلف روی گراف دارو و گراف وزندارکه برای پروتئین میسازد DTA را پیش بینی میکند. | یک گراف وزنی برای پروتئین ساخته شده که می تواند به طور جامع اطلاعات ساختار پروتئین را شرح دهد و به جای مقدار باینری که در DGraphDTA استفاده می شود، از احتمال تعامل به عنوان وزن یال های گراف استفاده می شود.از (GCN) و (GAT) برای استخراج بردارهای نهفته گراف های پروتئین و دارو، و پیش‌بینی DTAبر اساس بردارهای نهفته استفاده میکند. |
| DeepH-DTA  [2020] |  |  |
| DeepNC  [2022] |  |  |
| S2GC  [2022] |  |  |
| Graph-ATT-DTA  [2022] |  |  |
| FusionDTA  [2022] | مدل مبتنی بر یادگیری عمیق با استفاده از BiLSTM و multi-head linear attention  DTAرا پیش بینی میکند. |  |
| Multiscale-DTA  [2022] | از multi-scale CNNs ومکانیسم توجه برای پیش بینی DTAاستفاده میکند. | ابتدا رشته های دارو و پروتئین با روشهایEmbedding کد گذاری میشوند . سپس با عبور از دو بلوک CNN سه لایه و مکانیسم توجه ویژگی های چند مقیاسی و جامع مولکول های دارو و هدف استخراج میشوند وبعد از ترکیب با هم وارد لایه های FCمیشوند تا DTAرا پیش بینی کند. |
| GSAT-DTA  [2022] | از ترنسفورمر برای پروتئین و BiGRU وGNN برای دارو استفاده کرده است. | Graph–sequence attention به شبکه اجازه میدهد تا روی مرتبط ترین بخش داده های ورودی تمرکز کند و دقیقترین اطلاعات و ویژگیها را از رشته های SMILES و گراف دارو استخراج کند.  multi-head self-attention در ترنسفومر با استفاده از انکدر و دیکدر ویژگیهای رشته های پروتئین را از موقعیت های مختلف رشته استخراج میکند و در نهایت خروجی DTAتخمین زده میشود . |

**فصل4**

**نتیجه گیری**

کاربرد یادگیری عمیق در کشف دارو شامل مدل ها و الگوریتم های پیش بینی است که می تواند منجر به کاهش هزینه و زمان این فرایند شود. استفاده مجدد از دارو، فرآیند شناسایی کاربردهای جدید برای داروهای موجود بر روی بیماری های جدید است. این یک رویکرد مقرون به صرفه و به صرفه برای کشف و توسعه دارو است.

پیش‌بینی ارتباط دارو-هدف (DTA) گام مهمی در فرایند کشف دارو است که می‌تواند به تسریع این روند کمک قابل توجهی کند. در سالهای اخیر از شبکه‌ عصبی گراف (GNN) برای پیش‌بینی DTA استفاده شده است. داده های غیر اقلیدسی مانند ساختار مولکولهای دارو و ساختارهای پروتئین را می توان به طور موثر با استفاده از گراف نشان داد. GNNامکان مطالعه تعاملات و تداخلات دارویی و کشف آنتی‌بیوتیک‌های جدید را فراهم می‌کند. شناسایی تداخلات دارو-هدف می تواند فرآیند کشف دارو را سرعت ببخشد و در نتیجه توسعه دارو را تسهیل کند. GNN می تواند فعل و انفعالات پیچیده بین داروها و اهداف را با مدل سازی آنها به عنوان گراف به تصویر بکشد.

طبق مطالعات انجام شده به منظور بهبود دقت پیش بینی DTA ، انتخاب یک مدل GNN مناسب در مرحله اول بسیار مهم است. شبکه‌های کانولوشن گراف (GCN) و شبکه‌ های توجه گراف (GAT) دو مدل GNN محبوب هستند که برای پیش‌بینی ِDTA شده‌اند. دو معماری GCN و GAT و تعداد لایه های مختلفی که برای آنها در نظر گرفته میشود از پرکاربردترین معماری ها هستند که به نتایج خوبی هم دست یافته اند......

**مراجع**

[1]Chen, Wei, et al. "Artificial intelligence for drug discovery: Resources, methods, and applications." Molecular Therapy-Nucleic Acids (2023).

[2] Thafar, Maha, et al. "Comparison study of computational prediction tools for drug-target binding affinities." Frontiers in chemistry 7 (2019): 782.

[3] Kim, Jintae, et al. "Comprehensive survey of recent drug discovery using deep learning." International Journal of Molecular Sciences 22.18 (2021): 9983.

#####################

#####################

#####################

[1] Pan, Xiaoqin, et al. "Deep learning for drug repurposing: Methods, databases, and applications." Wiley interdisciplinary reviews: Computational molecular science 12.4 (2022): e1597..

[2] Yeu, Yunku, Youngmi Yoon, and Sanghyun Park. "Protein localization vector propagation: a method for improving the accuracy of drug repositioning." Molecular BioSystems 11.7 (2015): 2096-2102.

[3] Weininger, David. "SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules." Journal of chemical information and computer sciences 28.1 (1988): 31-36.

[4]: Li, Hongjian, et al. "Low-quality structural and interaction data improves binding affinity prediction via random forest." Molecules 20.6 (2015): 10947-10962.

[5]: Pahikkala, Tapio, et al. "Toward more realistic drug–target interaction predictions." Briefings in bioinformatics 16.2 (2015): 325-337.

[6]: He, Tong, et al. "SimBoost: a read-across approach for predicting drug–target binding affinities using gradient boosting machines." Journal of cheminformatics 9.1 (2017): 1-14.

[7]: Öztürk, Hakime, Arzucan Özgür, and Elif Ozkirimli. "DeepDTA: deep drug–target binding affinity prediction." Bioinformatics 34.17 (2018): i821-i829.

[8]: Öztürk, Hakime, Elif Ozkirimli, and Arzucan Özgür. "WideDTA: prediction of drug-target binding affinity." arXiv preprint arXiv:1902.04166 (2019).

[9]: Zhao, Lingling, et al. "GANsDTA: Predicting drug-target binding affinity using GANs." Frontiers in genetics 10 (2020): 1243.

[10]: Lin, Xuan. "DeepGS: Deep representation learning of graphs and sequences for drug-target binding affinity prediction." arXiv preprint arXiv:2003.13902 (2020).

[11]: Karimi, Mostafa, et al. "DeepAffinity: interpretable deep learning of compound–protein affinity through unified recurrent and convolutional neural networks." Bioinformatics 35.18 (2019): 3329-3338.

[13]: Nguyen, Thin, et al. "GraphDTA: predicting drug–target binding affinity with graph neural networks." Bioinformatics 37.8 (2021): 1140-1147.

[13]: Jiang, Mingjian, et al. "Drug–target affinity prediction using graph neural network and contact maps." RSC advances 10.35 (2020): 20701-20712.

[14]: Jiang, Mingjian, et al. "Sequence-based drug-target affinity prediction using weighted graph neural networks." BMC genomics 23.1 (2022): 1-17

1. Drug discovery [↑](#footnote-ref-1)
2. Artificial Intelligence [↑](#footnote-ref-2)
3. Drug-Target Interaction [↑](#footnote-ref-3)
4. Drug-Target binding affinity [↑](#footnote-ref-4)
5. Sequence-based method [↑](#footnote-ref-5)
6. Structure-based method [↑](#footnote-ref-6)
7. Simplified Molecular Input Line Entry System [↑](#footnote-ref-7)
8. Multi-Layer Perceptron [↑](#footnote-ref-8)
9. Convolutional Neural Network [↑](#footnote-ref-9)
10. Long Short Term Memory [↑](#footnote-ref-10)
11. Recurrent Neural Network [↑](#footnote-ref-11)
12. Generative Adversarial Network [↑](#footnote-ref-12)
13. Autoencoder [↑](#footnote-ref-13)
14. Graph Neural Network [↑](#footnote-ref-14)
15. Attention-Based Model [↑](#footnote-ref-15)
16. Random forest [↑](#footnote-ref-16)
17. Convolutional Neural Network [↑](#footnote-ref-17)
18. semi-supervised [↑](#footnote-ref-18)
19. Generative adversarial networks [↑](#footnote-ref-19)
20. Graph Convolutional Network [↑](#footnote-ref-20)
21. Graph Attention Network [↑](#footnote-ref-21)